RECEIVED
CENTRAL FAX CENTER
OCT 19 2004

IBM.5202

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	**Baumgartner *et al.***	Group Art Unit:	**2654**
Serial No.:	**09/498234**	Examiner:	**A. Armstrong**
Filed:	**02/03/2000**	Attorney Docket:	**AUS990879US1**
For:	**METHOD AND SYSTEM OF AUDIO FILE SEARCHING**		

I, the undersigned Joseph P. Lally, hereby certify that this correspondence is being facsimile transmitted to the USPTO or deposited with the US Postal Service with sufficient postage as first class mail in an envelope addressed to: MAIL STOP AMENDMENT, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

October 19, 2004
Date

Signature

AFFIDAVIT UNDER 37 CFR § 1.131

MAIL STOP AMENDMENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

I, the undersigned inventor of the above referenced application, having been admonished that willful false statements and the like are punishable by fine, imprisonment, or both (18 U.S.C. § 1001) and may jeopardize the validity of the application or any patent issuing thereon, declare as follows:

I am an inventor of the above captioned patent application [the Application]. As indicated in the document attached hereto as Exhibit "A" entitled *Disclosure AUS8-1999-1488 (Method and System for Audio File Searching Using Voice / Text Keys)* [the Disclosure], my co-inventors and I conceived of system for receiving a text-based input, converting the text input to a corresponding diphthong sequence, encoding the diphthong sequence, and using the encoded diphthong sequence to search and compare encodings of diphthong sequences taken from the audio content of a storage device such as a CD. As described in the Disclosure, the invention was workable at least as early as September 28, 1998. The invention was documented via the Disclosure on or about September 30, 1998, submitted to a patent review committee, sent to

outside counsel on or about November 19, 1999 and ultimately drafted and filed as the currently pending patent application on February 3, 2000.

I further declare that all statements made of my own knowledge are true and all statements made on information and belief are believed to be true.

Jason Baumgartner Date 10/19/2004

Nadeem Malik Date

Steven Roberts Date

Serial No. 09/498,234
Art Unit: 2654
Examiner. A. Armstrong
Docket: AUS990879US1

outside counsel on or about November 19, 1999 and ultimately drafted and filed as the currently pending patent application on February 3, 2000.

I further declare that all statements made of my own knowledge are true and all statements made on information and belief are believed to be true.

Jason Baumgartner	Date
Nadeem Malik	10/19/2004 Date
Steven Roberts	Date

Serial No. 09/498,234
Art Unit: 3654
Examiner. A. Armstrong
Docket: AUS990879US1

outside counsel on or about November 19, 1999 and ultimately drafted and filed as the currently pending patent application on February 3, 2000.

I further declare that all statements made of my own knowledge are true and all statements made on information and belief are believed to be true.

Jason Baumgartner — Date

Nadeem Malik — Date

Steven Roberts — Date 10/18/2004

EXHIBIT A

Disclosure AUS8-1999-1488

Created By: Avi Saha **Created On:** 09/30/98 02:23:55 PM
Last Modified By: Cheryl Work **Last Modified On:** 11/08/99 09:41:21 AM

*** IBM Confidential ***

Required fields are marked with the asterisk (*) and must be filled in to complete the form .

Summary

Status	Under Evaluation
Processing Location	AUS
Functional Area	07 - STQ - CORPORATE SOFTWARE TECH. (C. LOGAN)
Attorney/Patent Professional	Mark McBurney/Austin/IBM
IDT Team	Tim Dietz/Austin/IBM; Nadeem Malik/Austin/IBM
Submitted Date	11/05/99 10:51:38 PM
Owning Division	CHQ
PVT Score	To calculate a PVT score, use the 'Calculate PVT' button.
Lab	
Technology Code	
Incentive Program	(INC4) PC Server and Consumer Products

Inventors with Lotus Notes IDs

Inventors: Steven Roberts/Austin/IBM, Jason Baumgartner/Austin/IBM, Nadeem Malik/Austin/IBM

Inventor Name > denotes primary contact	Inventor Serial	Div/Dept	Manager Serial	Manager Name
> Roberts, Steven L.	785258	7T/OQ6A	169449	Rodriguez, Victor M.
Baumgartner, Jason	913486	7T/OQ6A	169449	Rodriguez, Victor M.
Malik, N. (Nadeem) Dr.	530606	47/SOKA	117025	Logan, Carol Angela

Inventors without Lotus Notes IDs

IDT Selection

IDT Team:	Attorney/Patent Professional:
Tim Dietz/Austin/IBM Nadeem Malik/Austin/IBM	Mark McBurney/Austin/IBM

Response Due to IP&L : 12/08/99

Main Idea

***Title of disclosure (in English)**

Method and System for Audio File Searching Using Voice / Text Keys

***Idea of disclosure**

1. Describe your invention, stating the problem solved (if appropriate), and indicating the advantages of using the invention.

This invention discloses a system for allowing the searching of audio files for vocal segments -- e.g., for a certain word or sequence of words. The search basis can be presented to the system in the form of a spoken word or typed equiva

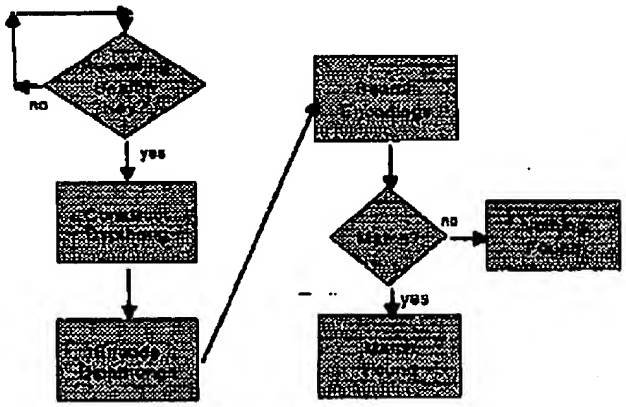
For instance, instead of fast forwarding to a portion of your favorite song on a compact disc or segment of a movie (with perhaps many backward-and-forward iterations), such a scheme would allow the user to speak (or type) the desired phrase (e.g., a song lyric, or line from the movie) and have the media player find the location of the desired item. Disclosed is a scheme that illustrates how such a system may be implemented.

Such a scheme has many advantages over existing state-of-the-art media devices, which require the user to reactively respond to the media content flashing across a television screen or an audio stream played at a such a rate that it is almost unintelligible (namely, manual high-speed fast-forward & rewind searching). This is a time consuming and somewhat annoying process. Using this disclosure, we can shift this tedious manual effort to the media device by building sufficient intelligence to allow the user to specify what they are looking for, and having the media device do the necessary scanning to find matches. Thus the user is thereby only presented with similar matches, and does not need to spend his/her own time interactively with the media device to find such matches. Such a search may further typically be carried out in a fast manner, since for example a 40x CD reader can parse audio from a CD at 40x the normal rate. A computer may further scan the high-speed analog signals from an analog video or audio media with higher precision than a human being, allowing for scanning of all forms of media (analog or digital) at higher speeds than achievable by a human, and without the annoying back-and-forth analysis typically suffered via manual and error-prone scanning.

2. How does the invention solve the problem or achieve an advantage,(a description of "the invention", including figures inline as appropriate)?
This invention pertains to the use of a high-performance audio speech interpretation system to allow for automated scanning of audio files for speech patterns (words or sequences of words). The user inputs the search key -- either by speaking the fragment to be searched for, or by entering text of said fragment. The system converts this key to a series of diphthongs (a primitive construct of the language of the audio file) using existing voice recognition technology -- these are represented as a string of symbols. The system next begins parsing the audio file -- this may be a digital file (e.g., a .wav file, a CD, or the audio track of a DVD), or an analog medium (e.g., an audio tape, or the audio track of a VCR tape) through an analog-to-digital converter. The media is transformed into a sequence of diphthongs -- again a string of symbols. As such, our symbolic audio file search is reduced to a string pattern matching.

A flowchart for this scheme is depicted in the figure below. First, the search pattern (entered manually by speech or
text) is encoded as a sequence of diphthongs. Next, the audio file to be searched is translated on-the-fly into a sequence of diphthongs. The "match" box utilizes standard pattern matching algorithms to look for instances of the search sequence within the file sequence. If a match is found, it is reported to the user (by playing the media from the point of the match). If this is not the correct match, the user may opt to "find next instance".

Voice Search on Dictionary Compressed Media



An alternate embodiment of this disclosure would also allow for "fuzzy pattern matching", which is useful for the followi reasons: 1) to desensitize the system to varients in speech-to-diphthong technology;
2) to allow the user use partial phrases.
Algorithms that do fuzzy pattern matches are in common use today (e.g., "suggestion" generators for spelling checker As the technology advances, more exact matches will result. This idea allows the development of the system without exposure to the exact behavior of the speech to diphthong technology.

An embodiment allowing partial phrases as input would permit the user to specify wildcards to further narrow down the fuzzy search. For example, assume that the user is really looking for the quote "All work and no play makes Jack a dul but does not recall the exact pattern, only that it starts with "All work" and ends with "dull boy". They can specify a wild between these fragments (to avoid too many matches, if either fragment was searched context-freely). The algorithm can optionally use some heuristics (or user-specified parameters) to limit the depth of the wildcard -- i.e., it is unlikely that there will be more than a few dozen diphthongs between the known fragments, and this parameter can av a match with 30 minutes of text between the fragments.

Our proposed solution to the implementation of this invention involves the use of a fast speech recognizor (e.g., a speech pattern-to-diphthong converter). Optionally, if text-based search pattern specification is to be carried out, a text-to-diphthong converter will be needed. The latter two are relatively straight-forward; speech recognition software exists today, and may be used for recognition of diphthongs. Similarly, a heuristic text-to-diphthong converter, which may be made exact by a dictionary file, can be employed using similar technology to those employed by "phone in and have your email read to you" systems.

This disclosed system has several advantages and applications. First is for home use in multimedia devices -- this scheme may greatly reduce the amount of time and manual error-prone effort involved in fast-forward and rewind based searching for scenes of a movie or parts of musical works. Second is in more technical fields, which may be used to search long narratives, interviews, proceedings, or surveillance files for exact phrases.

Note that such a scheme may be much faster than a human. Even barring the error-prone back-and-forth narrowing down of the target, a human can only comprehend up to a certain speed -- for example, if listening to an audio file, a human may only comprehend speech up to a factor of 8 or so. A computer may parse audio much faster -- either from a
digital source (e.g., a 40x CD ROM reader), or from an analog source (e.g., playback of an audio tape at high speed), and extract diphthongs from these high-speed sources. Such a scheme could also be

enhanced to be able to directly parse compressed audio files (e.g., MP3).

3. If the same advantage or problem has been identified by others (inside/outside IBM), how have those others solved it and does your solution differ and why is it better?
Other companies have successfully marketed this concept in forms like personal address assistants that will look up a person's phone number if you speak out the person's name. Our solution is more robust and can handle much larger jobs because of the flexibility gained by performing the diphthong analysis. Thus, if the dictionary can be searched in some logical fashion without being expanded, it is possible to apply the technique directly to compressed audio streams.

For instance, by modifying our example algorithm it is possible to find a diphthong stream in an MP3 file. In other words, this technique is superior because it works in an intermediate format that takes advantage of scaleable media whereas the existing techniques perform simple correlations on speech patterns over a time interval.

4. If the invention is implemented in a product or prototype, include technical details, purpose, disclosure details to others and the date of that implementation.

***Critical Questions (Questions 1 - 7 must be answered)**

***Question 1** On what date was the invention workable? 09/30/98 Please format the date as MM/DD/YYYY (Workable means i.e. when you know that your design will solve the problem)

***Question 2** Is there any planned or actual publication or disclosure of your invention to anyone outside IBM?	○ Yes ● No
If yes, Enter the name of each publication or patent and the date published below. Publication/Patent: Date Published or Issued:	
Are you aware of any publications, products or patents that relate to this invention?	○ Yes ● No
If yes, Enter the name of each publication or patent and the date published below. Publication/Patent: Date Published or Issued:	

***Question 3** Has the subject matter of the invention or a product incorporating the invention been sold, used internally in manufacturing, announced for sale, or included in a proposal?	○ Yes ● No
Is a sale, use in manufacturing, product announcement, or proposal planned?	○ Yes ● No
If Yes, identify the product if known and indicate the date or planned date of sale, announcements, or proposal and to whom the sale, announcement or proposal has been or will be made. Product: Version/Release: Code Name: Date: To Whom: If more than one, use cut and paste and append as necessary in the field provided.	

Method and System for Audio File S... ...ting Using Voice / Text Keys - continued

*Question 4	Yes / No
Was the subject matter of your invention or a product incorporating your invention used in public, e.g., outside IBM or in the presence of non-IBMers?	
If yes, give a date: **Please format the date as MM/DD/YYYY**	

*Question 5	Yes / No
Have you ever discussed your invention with others not employed at IBM?	
If yes, identify individuals and date discussed. Fill in the text area with the following information, the names of the individuals, the employer, date discussed, under CDA, and CDA #.	

*Question 6	Yes / No / Not sure
Was the invention, in any way, started or developed under a government contract or project?	
If Yes, enter the contract number	

*Question 7	Yes / No / Not Sure
Was the invention made in the course of any alliance, joint development or other contract activities?	
If Yes, enter the following: Name of Alliance, Contractor or Joint Developer	
Contract ID number	
Relationship contact name	
Relationship contact E-mail	
Relationship contact phone	

Question 8	Yes / No
Have you submitted, or are you aware of, any related disclosure submission?	
If Yes, please provide the title and docket or disclosure number below:	

Question 9

What type of companies do you expect to compete with inventions of this type? *Check all that apply.*

- ☐ Manufacturers of enterprise servers
- ☐ Manufacturers of entry servers
- ☐ Manufacturers of workstations
- ☒ Manufacturers of PC's
- ☒ Non-computer manufacturers
- ☒ Developers of operating systems
- ☒ Developers of networking software
- ☒ Developers of application software
- ☒ Integrated solution providers
- ☒ Service providers
- ☐ Other (Please specify below)

Patent Value Tool (Optional - this may be used by the inventor and attorney to assist with the eval

(The Patent Value tool can be used by you or the evaluation team to determine the potential licensing

value of your invention.)

The **Patent Value Tool** has not yet been used to calculate a score.

Post Disclosure Text & Drawings
Enter any additional information relating to this disclosure below:

(Form Revised 12/17/97)